

# Две истории биоинформатики: наука о данных vs наука о жизни

Михаил Волошин

Магистрант, философский факультет, Московский государственный университет им. М. В. Ломоносова (МГУ). Адрес: 119991, Москва, Ломоносовский пр-т, 27, корп. 4. E-mail: allrour95@gmail.com.

*Ключевые слова:* биоинформатика; Полина Хогевер; Халлам Стивенс; информация; данные; история науки; научная дисциплина.

В статье рассматриваются две реконструкции истории одной научной дисциплины — биоинформатики. Одна представлена автором термина «биоинформатика» Полиной Хогевер, другая — историком и социологом науки Халламом Стивенсом. Несмотря на личную вовлеченность Хогевер в становление биоинформатики как научной дисциплины, с одной стороны, и опору Стивенса на значительный объем микросоциологических, наукометрических и других свидетельств, с другой стороны, авторы рассказывают принципиально разные истории. По Хогевер, возникновение биоинформатики связано с новыми эпистемологическими запросами наук о жизни после ключевых открытий молекулярной биологии в середине XX века. Для Стивенса новая дисциплина стала результатом внедрения вычислительных методов и технологий из более математизированных областей в биологию. Это различие порождается расхождением в интерпретации того, что такое биоинформатика, которое, в свою очередь, зависит от предварительных онтологических допущений о природе живого.

Связь концепции жизни с концепцией информации у Хогевер объясняется с привлечением системного подхода; Стивенс, отвергая эту связь, акцентирует внимание на переносе научных практик из других дисциплин и порождаемых этим переносом новых способах видения живого. При этом попытка определить специфику биоинформатики как дисциплины обрекает обоих авторов на порочный круг: дисциплина определяется тем, что определяет сама, а именно некоторым представлением об информации и/или данных. Порочный круг обусловлен тем, что нормативный характер критерия разграничения дисциплин не позволяет адекватно объяснить возникновение каждой из них в отдельности. Взамен предлагается использовать дескриптивный критерий, понимаемый как исследование условий возможности уже имеющей место дифференциации научных практик. Определенное представление об информации/данных и связанная с ним онтология должны быть выявлены в результате исследования такого рода, а не предпосылаться ему.

## Что значит быть научной дисциплиной?

**З**АДАВАЯСЬ вопросом о специфике той или иной научной дисциплины — о том, что ее отделяет от других дисциплин и делает самостоятельной, — разумно вначале подумать о том, какой тип ответа будет для нас удовлетворительным, то есть какого рода критерий вообще используется для разграничения дисциплин. Широко известны два основных варианта, выработанных главным образом в ходе попытки определения специфики гуманитарного знания: специфическая предметная область и специфический методологический инструментарий. Отдавая, например, предпочтение первому варианту, мы будем считать биологию самостоятельной наукой — постольку, поскольку признаем наличие более-менее определенной границы между живой и неживой природой; в то же время положение химии окажется под вопросом — постольку, поскольку мы признаем редуцируемость химических процессов к физическим<sup>1</sup>. Тогда деление наук по предметным областям оказывается зависящим от некоего заранее имеющегося деления на предметные области, то есть от системы онтологических допущений, и кажется невероятным, чтобы такая система существовала до и независимо от соответствующих дисциплин. Тот же вопрос о границе неживого и живого сам по себе относится к предметной области биологии. Аналогичная проблема возникает, если использовать методологический критерий: специфические методы научной дисциплины часто вырабатываются в ее собственных рамках. Можно вспомнить, что сам методологический критерий возникает у неокантианцев Баденской школы как попытка обоснования независимого статуса гуманитарных наук, порождая своего рода замкнутый круг: 1) научная дисциплина имеет собственные методы и поэтому самостоятельна; 2) научная дисциплина самостоятельна и поэтому имеет право на собственные методы<sup>2</sup>.

1. В наши намерения не входит отстаивать ни одну из озвученных позиций — они носят здесь исключительно иллюстративный характер.
2. Заметим в скобках, что попытка выйти из этого круга превращает методологический критерий в предметный: например, можно сказать, что

При такой постановке проблемы выбор предметного или методологического критерия во многом оказывается произвольным, и любая полемика на этот счет становится бессмысленной. Тем не менее полемика ведется, и значит, некоторое основание выбора имплицитно присутствует. Это данность — или, если угодно, «эмпирический факт», — что существует нечто, называемое, допустим, биологией, и способ существования этого нечто сходен со способом существования научного знания. Сам этот «эмпирический факт» основан на том представлении, что для научного знания существовать — значит быть некоторым образом оформленным, будь то в виде организаций, институтов, учебников, научных журналов и т. д. Критерий самостоятельности дисциплины, таким образом, оказывается не только *нормативным* (как понять, что следует считать самостоятельной областью науки?), но и *дескриптивным* (науки уже как-то разграничены, но по какому принципу?).

Возможно, тогда в принципе не следует искать общие основания классификации, предпочтя социологию науки философии науки и признав сложившееся устройство научного знания результатом (во многом случайным) игры социальных отношений и серии (случайных) исторических событий. Научная область могла бы не возникнуть, если бы один ученый не встретился с другим, или если бы ему не удалось получить финансирование для своего проекта, или если бы не началась война. Это до определенной степени верно, но тем интереснее ситуация, когда в эмпирически фиксируемой и выделяемой области знания удается обнаружить не только социологические, но и концептуальные (предметные либо методологические) основания ее самостоятельности; когда серия случайностей предстает последовательно разворачивающейся *историей* развития науки.

Перед нами две попытки реконструкции истории одной дисциплины.

## **История термина: наука о жизни**

История термина «биоинформатика» восходит к 1970 году, когда была опубликована статья голландского биолога Полины Хогевер

гуманитарные науки имеют право на особые методы в силу их специфической соотнесенности с особым онтологическим регионом ценностей.

под названием «Биоинформатика: рабочее понятие»<sup>3</sup>. По признанию самой Хогевер, эта статья осталась незамеченной по причине нераспространенности голландского языка, и обычно ссылались на более позднюю статью на английском языке (1978)<sup>4</sup>.

Хогевер сделала попытку определить биоинформатику именно предметным образом: это «изучение информационных процессов в биотических системах»<sup>5</sup>. Поясняя это определение, она указывает на то, что с возникновением и развитием ряда концепций молекулярной биологии связь между жизнью и информацией становится очевидной. Таковы представления о генетическом коде, центральная догма молекулярной биологии как тезис об односторонности информационного потока (от ДНК к РНК и белку) и в целом «информационно-центричная» терминология молекулярной биологии. Биоинформатика должна была заняться распространением информационного подхода на различные уровни организации живого — от генотипа до взаимодействия с окружающей средой. Хогевер с сожалением констатирует, что развитие биоинформатики пошло по другому пути. Смещение акцентов со специфики процессов на разработку методов анализа больших данных произошло с открытием методов секвенирования, позволивших, по выражению отечественного биоинформатика Михаила Гельфанда, «получать факты индустриально»<sup>6</sup>. Поставив производство данных на поток, ученые превратили биоинформатику в *data-driven*-науку — исследовательскую область, озабоченную созданием баз данных, алгоритмов секвенирования, выравнивания последовательностей, то есть скорее специализированную область *Computer Science*, чем новый подход к биологии. Хогевер пишет, что и она сама, и те, кто пытался апеллировать к «первоначальному смыслу» биоинформатики, активно включились в работу в этой области.

Но в то же время мы упорно твердили, что биоинформатика — это не про разработку методов<sup>7</sup>.

3. Hesper B., Hogeweg P. Bioinformatica: een werkconcept // Kameleon. 1970. Vol. 1. № 6. P. 28–29.
4. Hogeweg P. Simulating the Growth of Cellular Forms // Simulation. 1978. Vol. 31. P. 90–96.
5. *Idem*. The Roots of Bioinformatics in Theoretical Biology // PloS Computational Biology. 2011. Vol. 7. № 3.
6. Гельфанд М. Н. Что может биоинформатика // Химия и жизнь. 2009. № 9.
7. Hogeweg P. The Roots of Bioinformatics in Theoretical Biology.

Хогевег сама задается вопросом: не было ли ее первоначальное определение всего лишь

... историческим анекдотом, вытесненным более распространенным значением термина — разработка и использование вычислительных методов<sup>8</sup>.

Ее ответ заключается в том, что современные биологи (на 2011 год) все меньше готовы поддерживать идею, будто организмы — это своего рода «мешки с белками»: так она называет установку, согласно которой свойства организма как целого могут быть редуцированы к свойствам составляющих его макромолекул. На смену этой установке приходит осознание необходимости моделирования динамических самоорганизующихся систем, неизбежно включающих в себя процессы информационного характера (здесь Хогевег ссылается, например, на работы Людвига фон Берталанфи<sup>9</sup>). Таким образом, исходное понимание биоинформатики как изучения информационных процессов возрождается в XXI веке.

Можно увидеть на примере этой истории, насколько проблематичными кажутся автору термина «биоинформатика» оба корня этого слова. Чтобы определить специфику биоинформатики так, как это делает Хогевег, нужно принимать (на веру?) целый ряд допущений, среди которых будут существование такой вещи (вещи ли? феномена? процесса?), как информация, ее роль в «биотических» системах (вероятно, отличная от ее роли вне этих систем), возможность работы с информацией компьютерными методами и т. д. Для Хогевег эти допущения — онтологические: *существуют* некие информационные процессы, которые надо изучать, и биоинформатика этим займется; более того, если она этим не займется, то мы впадем в редукционизм, а значит, не познаем истинную природу жизни, поскольку истинная природа жизни — системная, динамическая и информационная, и в наш просвещенный XXI век эта идея наконец начала проникать в умы.

Авторы исследования, посвященного главным образом этическим аспектам биоинформатики и проведенного в 2006 году в Институте Пастера (Лилль, Франция), указывают, что там существуют противоречивые интерпретации того, чем занимается биоин-

8. *Hogeweg P. The Roots of Bioinformatics in Theoretical Biology.*

9. В частности, на: *Von Bertalanffy L. General System Theory.* N.Y.: George Braziller, 1973.

форматика. С одной стороны, официальный сайт организации гласит, что ученые-биоинформатики призваны обеспечить

...вычислительные мощности, хранилище данных, полученных исследователями группами, доступ к наиболее распространенному программному обеспечению, знания и компетенции инженеров в области обработки данных и биоинформатики, создание алгоритмов и программ, поддержку исследовательских групп в компьютерной лаборатории Лилля и учебные курсы по биоинформатике<sup>10</sup>.

С другой стороны, они цитируют французского ученого-биоинформатика Жана-Мишеля Клавери, который утверждал, что

...окончание *-informatics* следует понимать как указание на интерпретацию биологической «информации», а не на использование компьютера<sup>11</sup>.

По мнению Армель де Бове и соавторов, биоинформатика — это мощный инструментарий, который используется не вполне адекватным объекту исследования способом. Ссылаясь на работу Мартина Манера и Марио Бунге «Основания биофилософии»<sup>12</sup>, они пишут о двух возможных типах редукции в биологии: «микроредукции» — стратегии объяснения целого организма через анализ составляющих его частей (в биоинформатике это анализ генома, транскриптома, протеома<sup>13</sup> организма) — и «макроредукции», объясняющей организм как элемент более крупного целого, например популяции (биоинформатическим примером макроредукции Бове и коллеги считают исследования в области генетической эпидемиологии). Но в обоих случаях перед нами редукционизм, некоторым образом устранивающий непосредственный предмет научного интереса. Биоинформатика же открывает возможность «более холистского подхода к живому»: благодаря ей стала доступна «интерпретация процессов в живой природе в со-

10. Bouvet A. et al. Bioinformatics: The Philosophical and Ethical Issues at Stake in a New Modality of Research Practices // *Medicine, health care and philosophy*. 2006. Vol. 9. P. 201–209.

11. Ibid. P. 205.

12. Mahner M., Bunge A. *Foundations of Biophilosophy*. B.: Springer, 1997.

13. По аналогии с геномом транскриптом организма — совокупность молекул РНК организма, получающихся в результате транскрипции; протеом — аминокислотные последовательности экспрессируемых в данном организме белков.

ответствии с теориями сложности и нелинейной динамики»<sup>14</sup>. Здесь вновь повторяются мотивы, звучащие у Хогеверг: мы (ученые-биологи) поняли, что организмы не «мешки с белками», и наша биоинформатическая методология должна соответствовать этому; биоинформатика, конечно, является «анализом данных», но пафос и Бове, и Хогеверг состоит в том, что ей предстоит (снова) стать чем-то большим.

Обе статьи, однако, подталкивают читателя к выводу, прямо противоположному желаним их авторов. Нет никакого смысла призывать к изменению представлений о биоинформатике, если она действительно изучает динамические самоорганизующиеся системы; с другой стороны, если для изучения динамических самоорганизующихся систем будут нужны биоинформатики, то они этим займутся. Аргументация Хогеверг и Бове с коллегами, по сути, направлена не против ошибочных представлений о биоинформатике, а против ошибочных представлений *о жизни*. Откуда в таком случае авторам известна истина о жизни, если дисциплина, которая должна была предъявить нам эту истину, так и не стала тем, чем должна была стать? И что делать с «неправильной», «редукционистской» биоинформатикой, которая существовала все эти годы в качестве бурно развивающейся области? Что она такое?

## История практик: наука о данных

Альтернативную позицию занимает автор одной из немногих книг<sup>15</sup>, специально посвященных историко-научному и философско-научному анализу биоинформатики, — Халлам Стивенс. Свое изложение истории биоинформатики он называет *data-driven history* — именно в силу того, что оно основано на отслеживании истории способов, которыми производятся, хранятся, перерабатываются и распределяются биологические данные. Внимание Стивенса приковано к практикам работы с данными, и во многом его книга написана с позиций социологии науки, в том числе микросоциологического подхода — изучения лабораторных практик.

Точного определения понятия «данные» Стивенс не дает.

14. Bouvet A. et al. Bioinformatics. P. 206.

15. Stevens H. Life Out of Sequence: A Data-Driven History of Bioinformatics. Chicago: University of Chicago Press, 2013.

Данные могут быть числами, отрывками текста, буквами последовательности ДНК или частями изображений. Когда биологи используют компьютеры, они некоторым образом работают с данными. Эти данные не являются знанием. Биолог может отправить данные в базу данных, но он или она не может отправить их в журнал для публикации. Их ценность — в том, что... они могут быть использованы для производства знания<sup>16</sup>.

Отметим, что Стивенс разводит не только понятия «данные» и «знания», но и понятия «данные» и «информация». Это нужно ему, чтобы

... избежать предположения, будто существует нечто в теле организма (код, информация), что каким-то образом попадает в компьютер. Информация рассматривается как нематериальное сообщение, закодированное в химии ДНК, которое впоследствии извлекается в компьютерный код. Данные, напротив, не перемещаются свободно за пределы биологических тел или через них — они тесно привязаны к аппаратуре и программному обеспечению, в которых они существуют. Данные принадлежат только компьютерам<sup>17</sup>.

Нам сразу бросается в глаза фундаментальное расхождение с идеями автора термина «биоинформатика». Стивенс, разумеется, упоминает статью Хогевет (причем более позднюю, 1978 года) и признает за ней авторство термина, при этом само определение Хогевет в книге не встречается ни разу. Можно предположить, что для Стивенса — с точки зрения *data-driven history* — существование какой бы то ни было концептуальной позиции по этому вопросу не принципиально, пока эта позиция не эксплицирована в социологически фиксируемых научных *практиках*; из дальнейшего изложения истории биоинформатики становится ясно, что задача поиска такой концептуальной позиции вообще нерелевантна самому подходу *data-driven history*. Вот замечательный пример того, как оптика исследователя (в нашем случае — историка науки) формирует предметную область: *data-driven history of science* превращается в *history of data-driven science*.

Не следует думать, утверждает Стивенс, будто все дело в том, что компьютеры становились все лучше, пока, наконец, не достигли того уровня, который сделал возможным их примене-

16. Stevens H. Life Out of Sequence. P. 6.

17. Ibid. P. 7.

ние в биологии. Напротив, в самой биологии произошли изменения, позволившие ей стать компьютеризованной дисциплиной. Эти изменения шли в двух направлениях. Во-первых, биология должна была стать математизированной дисциплиной, а во-вторых, необходимо было задействовать статистические методы. Задачей биологии в этом плане представлялось воспроизведение пути, пройденного физическими науками. В XX веке и особенно в послевоенный период, как пишет Алвин Вайнберг, в физике произошла значительная трансформация способа институционализации науки — переход к так называемой *Big Science*, от отдельных лабораторий и индивидуальных ученых — к громадным коллективам лабораторий, от случайного и негарантированного финансирования — к систематической работе на государственный заказ, от самодельных приборов — к поставленному на поток производству сложного и дорогого научного оборудования<sup>18</sup>. Результатом этого было создание атомной энергетики, атомного оружия, космические полеты и т. д. Естественно, повторить успех физики — заманчивая затея. Евгений Кунин даже предложил по этому поводу в очередной раз перефразировать Феодосия Добржанского и заявил:

Ничто в эволюции — и в популяционной генетике — не имеет смысла, кроме как в свете статистической физики<sup>19</sup>.

Стивенс называет имена ответственных за компьютеризацию биологии по модели физики. Это прежде всего Роберт Лэдди, в 1965 году опубликовавший книгу «Использование компьютеров в биологии и медицине», в которой предложил целую серию возможных физико-математических подходов к решению биологических проблем. Это Эдвард Фейгенбаум и Джошуа Ледерберг, создатели *MOLGEN* — первой «базы знаний» по молекулярной биологии, включавшей в себя элементы искусственного интеллекта и задумывавшейся как программы в помощь ученому-экспериментатору. Это Уолтер Гоад, первоначально сотрудник известной лаборатории Лос-Аламоса, в которой он занимался компьютерным моделированием взрыва водородной бомбы, для чего ему пришлось использовать статистические методы (в том числе метод Монте-Карло), — и этот свой «статистический» взгляд на вещи он перенес

18. Weinberg A. M. Impact of Large-Scale Science on the United States // Science. 1961. Vol. 134. P. 161–164.

19. Кунин Е. Логика случая. О природе и происхождении биологической эволюции. М.: Центрполиграф, 2014. С. 452.

на биологические объекты: чтобы изучать биологию, необходимы огромные массивы данных, которые можно обрабатывать методами, уже продемонстрировавшими свою эффективность в физике. Результатом многолетней работы Гоада стал *GenBank* — самая известная в мире база данных последовательностей генов и геномов. Наконец, это Джеймс Остелл, который в 1979 году получил доступ к *MOLGEN* и остался недоволен полным отсутствием «дружественного интерфейса» у программ *MOLGEN*, а также тем, что их работа требовала огромных аппаратных ресурсов. Остелл разработал первый пакет программ для молекулярной биологии, который мог работать на микрокомпьютере (то есть компьютере приблизительно современных размеров, что означало возможность массового внедрения) и был притом *user-friendly* — понятен рядовому сотруднику лаборатории, не очень хорошо знакомому с программированием.

Таким образом, повсеместное внедрение компьютеров в биологию Стивенс склонен связывать с конкретными людьми, обладавшими конкретными интересами и навыками, а также с конкретными институциями. «Биологи начали использовать компьютеры, потому что стали доверять этим новым способам мыслить и действовать», — пишет Стивенс, и это доверие стало возможным благодаря «переориентации институтов и практик, переструктурированию способов производства биологического знания»<sup>20</sup>. Однако, на наш взгляд, эта позиция не свободна от противоречий. С одной стороны, распространенное применение вычислительной техники делает возможным массовое накопление данных (и вообще существование этих данных — в соответствии с тезисом о том, что скорее биология подстраивалась под компьютерные технологии, чем наоборот). С другой — необходимость в применении компьютеров обосновывается тем, что данных *уже* слишком много. Выходит, что проблема была лишь в отсутствии определенных людей, способных представить эти данные таким образом, чтобы компьютеры можно было применять. Так что указание Стивенса на конкретные пути взаимодействия биологии с вычислительной техникой, по нашему мнению, — хорошая иллюстрация, но плохое объяснение. Извечный вопрос о «роли личности в истории» решается здесь довольно однозначно при полном отсутствии рефлексии по этому поводу: имплицитно должно предполагаться, что существуют выдающиеся личности, творческому порыву которых подчиняются целые сети социальных взаимо-

20. *Stevens H. Life Out of Sequence. P. 40–41.*

действий в науке, в то же время эксплицитно Стивенс не умеет демонстрировать свою близость микросоциологии и акторно-сетевому подходу, в рамках которых сеть в принципе не может полностью детерминироваться одним или несколькими «особенно выдающимися» акторами.

Применение компьютерных методов к биологии в описанный период 1960–1980-х годов не является, по Стивенсу, собственно биоинформатикой. Возникновение биоинформатики как дисциплины он относит к 1989–1992 годам в связи с резким (в 10 раз) скачком числа публикаций с ключевыми словами «биоинформатика», «вычислительная биология», «вычислительная И биология», наблюдавшимся в эти годы, а также с организацией ряда институтов — Национального центра биотехнологической информации (NCBI, 1988) и Европейского института биоинформатики (EBI, 1993). В это же время появляются первые университетские программы по биоинформатике (в 1993 году, в Университете Райса и Хьюстонском университете), а в 1995 году издается первый учебник по биоинформатике под авторством Майкла Уотермана<sup>21</sup>.

Необходимость в такой институционализации Стивенс вновь объясняет ростом количества данных. Данных не просто много — они растут неконтролируемо, из-за чего, в частности, ученые стали говорить об «информационном взрыве» и «наводнении данными». Стивенс пишет:

Использование компьютера в качестве повседневного лабораторного инструмента — это одно, но организация полномасштабной координации и сотрудничества требовала специализированных знаний и умений... Контролирование потоков данных, их фильтрация и перенаправление в нужные места — не тот же самый род деятельности, что и их повседневное использование, но столь же важно<sup>22</sup>.

Биоинформатика как дисциплина появляется, соответственно, в ответ на требование о заполнении лакуны: отсутствовали биологи, способные справиться с большими объемами данных, однако эту проблему на протяжении более полувека успешно решали физики и математики — слабо знакомые со спецификой данных, которые им предстояло обработать. Наконец, еще одним мощным фактором развития стал *Human Genome Project (HGP)* — огром-

21. Waterman M. S. Introduction to Computational Biology: Maps, Sequences and Genomes. Boca Raton: Chapman&Hall/CRC, 1995.

22. Stevens H. Life Out of Sequence. P. 35.

ный по своему замыслу проект, требовавший, по мере своей реализации, все более и более эффективных инструментов работы с данными. По мнению Стивенса, *HGP* — это уже в полной мере «большая наука», в том же смысле, который этому термину придавал Вайнберг. И не только по мнению Стивенса; Артур Леск, автор одного из самых известных учебников по биоинформатике, отмечает:

Расшифровка последовательности генома человека стоит в одном ряду с Манхэттенским проектом, который разрабатывал атомное оружие во время Второй мировой войны, и с освоением космоса. Эти проекты занимали и занимают лидирующие позиции как в фундаментальной науке, так и в широкомасштабном и дорогостоящем развитии прикладных исследований. Биология никогда не будет в таком выгодном положении, не будет она рассчитывать и таким бюджетом<sup>23</sup>.

Такой подход к определению исторического момента возникновения дисциплины выглядит намного более убедительным, чем отсылка к первому использованию термина. Но так ли уж велика разница между ними? Появляются статьи, научные организации, учебные программы, активно использующие термин «биоинформатика»; они являются реакцией на количественный рост данных. В то же время *типы* этих данных и, соответственно, *типы операций* с этими данными остаются прежними: они унаследованы из биологии, «компьютеризованной» по образцу физики до 1990-х годов<sup>24</sup>. Можно было бы защитить позицию Стивенса, заявив, что названия несущественны: в самом деле, зачем уделять столь значительное внимание тому, что в 1990-х годах дисциплина стала называться иначе, если смена названий (которая вряд ли могла

23. Леск А. Введение в биоинформатику. М.: Бином. Лаборатория знаний, 2009. С. 9.

24. Например, Джоэль Хаген указывает, что Джон Кендрю использовал компьютер для расчетов трехмерной структуры миоглобина еще в 1957 году, а построение филогенетических деревьев для гомологичных белков разных организмов (в частности, цитохрома С) было осуществлено Уолтером Фитчем и Эмануэлем Марголиашем в 1967 году. Таким образом, современные методы биоинформатики являются наследниками разработок докомпьютерной эпохи (то есть до того периода в истории биологии, который описан Стивенсом как *data-driven history*). Само по себе это не противоречит описанию истории биоинформатики Стивенсом, однако ставит под сомнение некоторые производимые им обобщения (*Hagen J. B. The Origins of Bioinformatics // Nature Reviews Genetics. 2000. Vol. 1. P. 231–236*).

быть мгновенной) отразила не более чем рост популярности биоинформатических методов. Однако сам Стивенс заявляет:

Такие ярлыки, как «биоинформатика», «вычислительная биология», «системная биология», значимы, потому что они обозначают специфические конфигурации и иерархии знания, доверия и власти в биологическом сообществе<sup>25</sup>.

Таким образом, количественный рост данных порождает новые конфигурации «знания, доверия и власти», которые, в свою очередь, порождают количественный рост данных, который порождает новые конфигурации «знания, доверия и власти»... Радиализация такого способа аргументации приводит нас к странной как минимум идее о том, что биоинформатика должна существовать вечно — иначе когда именно она все же должна была возникнуть?

Проблема истории Стивенса очень похожа на проблему истории Хогевег. Стивенс заранее должен иметь институты, формирующие оптику рассмотрения биологических данных как биоинформатических данных, но объясняет возникновение этих институтов новыми данными. Хогевег заранее должна иметь представление об «информационности» процессов в живых системах, чтобы обосновать необходимость биоинформатики, которая и должна производить такого рода представление. Кажется, что оба это понимают и пытаются подключить к своим историческим объяснениям «внешние» силы, разрывающие этот порочный круг: Хогевег говорит о заимствовании концептуального аппарата у общей теории систем, Стивенс — о переносе практик работы с данными из статистической физики. Этот, по-видимому, необходимый ход производится не без внутренних противоречий с собственными позициями обоих. Однако и сама возможность такого объяснения должна быть обоснована: *почему* мы считаем, что биологические объекты таковы, что их можно изучать с помощью общей теории систем? *Почему* перенос методов статистической физики на биологию оказывается успешным и продуктивным?

## Исследование условий возможности

Можно задаться вопросом о том, совместимы ли эти истории. В самом деле, Хогевег предложила — вместе с термином — неко-

25. Stevens H. Life Out of Sequence. P. 44.

торую исследовательскую программу в биологии, реализовать которую помогут компьютеры; Стивенс же констатирует, что биоинформатика не пошла по этому пути, и предлагает детальнее посмотреть на ту биоинформатику, которая все же имела место. В этом смысле обе истории верны — нужно лишь не считать, что существует «истинная» и «ложная» биоинформатика, а обратиться к «фактам самим по себе». Проблема в том, что подбор релевантных для той или иной истории фактов обусловлен — по крайней мере, в данном случае — не просто различными значениями слова «биоинформатика», а различными *онтологиями живого*, которые предпосылаются в качестве самоочевидных всем дальнейшим «историческим» рассуждениям.

1. Живое существует в форме динамических саморазвивающихся систем, в которых существенную роль играют информационные процессы; биологи могут изучать эти системы с помощью компьютеров.
2. Живое — это материальные объекты и процессы, не имеющие отношения к информации; биологи могут получать данные о них и обрабатывать эти данные на компьютерах, производя знание.

Это различие порождает разные трактовки причины возникновения новой научной дисциплины.

1. Биоинформатика возникает как попытка решить концептуальную проблему: как описывать, моделировать эти системы, предсказывать их поведение? Компьютеры предлагают новые способы научного объяснения.
2. Биоинформатика возникает как попытка решить техническую проблему: как обрабатывать такое огромное количество данных? Компьютеры предлагают новые вычислительные мощности (которые воплощаются в новые конфигурации «знания, доверия и власти»).

И при этом ни та ни другая трактовка не является удовлетворительной. Дисциплина, возникновение которой «объясняет» Хогеве, никогда не существовала — по крайней мере, под ярлыком «биоинформатики»<sup>26</sup>. Стивенс в своих «объяснениях» апеллирует то ли к акторно-сетевому подходу, то ли к дискурсу знания-

26. Отдельный вопрос, который мы не можем здесь обсудить подробно: а существуют ли в современной биологии холистские подходы, способные избежать обвинений в редукционизме, и возможны ли они вообще? Полемика по поводу понятия «информация» в философии биологии во многом инспирирована проблематикой генетического детерминизма и связанного с ним редукционистского взгляда на развитие организ-

власти, то ли вообще ко второму закону диалектики Гегеля, в конечном счете сводя возникновение дисциплины к энтузиазму отдельных ученых.

Различие в онтологических допущениях, на которых основываются обе истории, состоит главным образом в интерпретации понятия «информация». Мы выше приводили высказывание Клавери, акцентировавшего внимание на *-informatics* в слове *bioinformatics* и зафиксировавшего две основные коннотации этого корня: информация в живых системах и информация в компьютерах (которую Стивенс называет «данными»). Чтобы рассказать историю биоинформатики, и Стивенсу, и Хогеверг пришлось определиться с тем, что это такое. Возможно, стоило идти в обратном направлении? У нас есть достаточно различимый набор практик работы с данными, который институционально оформлен в дисциплину под названием «биоинформатика». Понятие информации — центральное для используемого в ней концептуального аппарата. Поставим вопрос иначе: а что именно должно подразумеваться под информацией, чтобы была вообще возможна такая дисциплина, как биоинформатика? Чем должна быть информация, чтобы методы, разрабатываемые биоинформатикой, могли претендовать на способность производить достоверное знание?

Хотя такая постановка вопроса отдаленно напоминает кантовскую идею критики, в ней гораздо больше общего с методологией, описанной Мишелем Фуко в «Порядке дискурса». Дисциплина, говорит Фуко,

...определяется областью объектов, совокупностью методов и корпусом положений, которые признаются истинными, равно как и действием правил и определений, техник и инструментов<sup>27</sup>.

Мы можем описать эту область как совокупность процедур контроля и ограничения дискурса (как это пытается делать Стивенс), избавив себя от иллюзий вроде фигуры автора или истины (присущих Хогеверг). Этого, однако, недостаточно: применив первый из четырех принципов метода Фуко (принцип переворачивания), следует затем привести в действие и все остальные (принципы

ма. См., напр.: *Griffiths P. E. Genetic Information: A Metaphor in Search of a Theory // Philosophy of Science. 2001. Vol. 68. P. 394–412.*

27. Фуко М. Порядок дискурса // Фуко М. Воля к истине: по ту сторону знания, власти и сексуальности. Работы разных лет. М.: Касталь, 1996. С. 65.

прерывности, специфичности и внешнего). Нас интересует последний, четвертый принцип, или правило внешнего:

...идти не от дискурса к его внутреннему и скрытому ядру, к некоей сердцевине мысли или значения, якобы в нем проявляющихся, но, беря за исходную точку сам дискурс, его появление и его регулярность, идти к внешним условиям его возможности, к тому, что дает место для случайной серии этих событий и что фиксирует их границы<sup>28</sup>.

Понятие информации (или данных) в этом смысле не должно для нас являться «внутренним и скрытым ядром» биоинформатики, оно не должно полагаться как имеющее заранее предзаданное (кем, интересно?) значение, конституирующее биоинформатический дискурс. В то же время это не значит, что оно не важно: на примере двух историй биоинформатики мы видим, что оно может функционировать как механизм спецификации дисциплины, ее самоопределения. «Значением» этого понятия и является этот механизм; ответ на вопрос о роли информации в живых системах или о способах организации данных в вычислительных машинах зависит от того, где этот механизм проводит границу между дисциплинами, а не наоборот. Задача исследователя в этом плане — выйти к границе с тем, что биоинформатикой не является, и обнаружить, как она проведена.

Возвращаясь к тому, что критерий самостоятельности дисциплин вообще является не только нормативным, но и дескриптивным, можно теперь уточнить этот тезис. Понятно, что «самостоятельность» любой дисциплины можно поставить под сомнение: наука не делится однозначным образом на изолированные области, как и сама не является четко выделенной из всего остального областью. Тем не менее она дисциплинарно организована, и то, что относится к одной из дисциплин — биоинформатике, — можно отличить от того, что к ней не относится, по ярлыку «биоинформатика», по сети взаимных цитирований в научных журналах, по совместной работе в исследовательских проектах и другим параметрам. Задаваясь вопросом о том, что выделяет биоинформатику (или что бы то ни было) из других научных дисциплин, мы можем указать на эти параметры, но тем самым мы ответим не на вопрос «Что выделяет?», а на вопрос «Как уже выделено?», что само по себе интересно, но ничего не объясняет. История био-

28. Фуко М. Порядок дискурса. С. 80.

информатики может быть серией случайных конфигураций «знания, доверия и власти», а не саморазвертыванием какой-нибудь идеи (например, идеи информации), но вопрос состоит в том, что отделило эту серию случайностей от других серий, что *сделало эти события возможными*. Deskриптивность критерия, соответственно, должна пониматься не как указание на фактическое наличие относительно независимых дисциплин, а как выявление самого способа проведения границ, реализованного в конкретном случае.

Такой критерий в итоге может оказаться предметным (например, если в результате исследования выяснится, что биоинформатика как дисциплина (само)определена работой со специфически понимаемой «информацией»), методологическим (например, если будет установлено, что методы, используемые биоинформатиками, чем-то принципиально отличаются от того, что делают физики или лингвисты) или каким бы то ни было еще. Существенно то, что, с одной стороны, установление этого критерия не может предварять исследование историка, философа или социолога науки, предписывая дисциплине определенную онтологию (это дело дисциплины, а не ее исследователя), а с другой — исследование не может ограничиться констатацией набора практик без попытки определить, что придает этим практикам дисциплинарное единство.

### *Библиография*

- Гельфанд М. Н. Что может биоинформатика // *Химия и жизнь*. 2009. № 9.
- Кунин Е. Логика случая. О природе и происхождении биологической эволюции. М.: Центрполиграф, 2014.
- Леск А. Введение в биоинформатику. М.: Бином. Лаборатория знаний, 2009.
- Фуко М. Порядок дискурса // Он же. Воля к истине: по ту сторону знания, власти и сексуальности. Работы разных лет. М.: Касталь, 1996. С. 47–96.
- Bouvet A., Deschamps C., Boitte P., Boury D. Bioinformatics: The Philosophical and Ethical Issues at Stake in a New Modality of Research Practices // *Medicine, health care and philosophy*. 2006. Vol. 9. P. 201–209.
- Griffiths P. E. Genetic Information: A Metaphor in Search of a Theory // *Philosophy of Science*. 2001. Vol. 68. P. 394–412.
- Hagen J. B. The Origins of Bioinformatics // *Nature Reviews Genetics*. 2000. Vol. 1. P. 231–236.
- Hesper B., Hogeweg P. Bioinformatica: een werkconcept // *Kameleon*. 1970. Vol. 1. № 6. P. 28–29.
- Hogeweg P. Simulating the Growth of Cellular Forms // *Simulation*. 1978. Vol. 31. P. 90–96.
- Hogeweg P. The Roots of Bioinformatics in Theoretical Biology // *PloS Computational Biology*. 2011. Vol. 7. № 3. e1002021.

- Mahner M., Bunge A. Foundations of Biophilosophy. B.: Springer, 1997.
- Stevens H. Life Out of Sequence: A Data-Driven History of Bioinformatics. Chicago: University of Chicago Press, 2013.
- Von Bertalanffy L. General System Theory. N.Y.: George Braziller, 1973.
- Waterman M. S. Introduction to Computational Biology: Maps, Sequences and Genomes. Boca Raton: Chapman&Hall/CRC, 1995.
- Weinberg A. M. Impact of Large-Scale Science on the United States // Science. 1961. Vol. 134. P. 161–164.

## TWO HISTORIES OF BIOINFORMATICS: DATA SCIENCE VS LIFE SCIENCE

MIKHAIL VOLOSHIN. MA student, Philosophy Faculty, allrour95@gmail.com.  
Lomonosov Moscow State University (MSU), 27 Lomonosovsky Ave., Bldg 4, 119991  
Moscow, Russia.

*Keywords:* bioinformatics; Paulien Hogeweg; Hallam Stevens; information; history of science; scientific discipline.

This paper focuses on two attempts to reconstruct the history of bioinformatics as a scientific discipline. Paulien Hogeweg who coined the term “bioinformatics,” presents one of them, while Hallam Stevens, who is both a historian and sociologist of science, offers the other. Although both of them can speak authoritatively about bioinformatics inasmuch as Hogeweg was personally involved in creating the field while Stevens has amassed a substantial amount of microsociological, scientometric and other evidence, they tell two fundamentally different stories. According to Hogeweg, bioinformatics came about as a response to new epistemic demands on the life sciences that arose from several key discoveries in molecular biology in the middle of 20th century. For Stevens, the new discipline was the result of transplanting computational methods and technologies into biology. This difference stems from divergent interpretations of what bioinformatics is, and these in turn depend upon different ontological claims about the nature of living things. The link between the concepts of life and information is explained by Hogeweg through a systems approach. Stevens discounts that link and concentrates instead on the transposition of scientific practices from other disciplines and on the new ways of understanding the living which are generated by this transposition.

The attempt to define bioinformatics as a scientific discipline ends for both of these theorists a tautology: the discipline is defined by something defined by this same discipline, that is, by a certain idea about information and/or data. The effect of this tautology is that a normative criterion for delimitation of disciplines (a set of requirements which are necessary and sufficient for considering a field of research as a scientific discipline) does not allow us to explain how each of them occurs individually. Instead, a descriptive criterion is proposed, which is to be understood as the study of the conditions which make possible the differentiations in scientific practices which have already taken place. A distinct understanding of information or data and the ontology associated with it should be the outcome of a study of this kind and not presupposed by it.

DOI: 10.22394/0869-5377-2020-3-1-18

### References

- Bouvet A., Deschamps C., Boitte P., Boury D. Bioinformatics: The Philosophical and Ethical Issues at Stake in a New Modality of Research Practices. *Medicine, Health Care and Philosophy*, 2006, vol. 9, pp. 201–209.
- Foucault M. Poriadok diskursa [L'ordre du discours]. *Volia k istine: po tu storonu znaniia, vlasti i seksual'nosti. Raboty raznykh let* [Will to Truth: Beyond Knowledge, Power, and Sexuality. Works of Various Years], Moscow, Kastal', 1996, pp. 47–96.
- Gelfand M. Chto mozhet bioinformatika [What Bioinformatics Can Do]. *Khimiya i Zhizn* [Chemistry and Life], 2009, no. 9.

- Griffiths P. E. Genetic Information: A Metaphor in Search of a Theory. *Philosophy of Science*, 2001, vol. 68, pp. 394–412.
- Hagen J. B. The Origins of Bioinformatics. *Nature Reviews Genetics*, 2000, vol. 1, pp. 231–236.
- Hesper B., Hogeweg P. Bioinformatica: een werkconcept. *Kameleon*, 1970, vol. 1, no. 6, pp. 28–29.
- Hogeweg P. Simulating the Growth of Cellular Forms. *Simulation*, 1978, vol. 31, pp. 90–96.
- Hogeweg P. The Roots of Bioinformatics in Theoretical Biology. *PLoS Computational Biology*, 2011, vol. 7, no. 3, e1002021.
- Koonin E. *Logika sluchaia. O prirode i proiskhozhdenii biologicheskoi evoliutsii* [The Logic of Chance. The Nature and Origin of Biological Evolution], Moscow, Tsentrpoligraf, 2014.
- Lesk A. *Vvedenie v bioinformatiku* [Introduction to Bioinformatics], Moscow, Binom. Laboratoriia znaniia, 2009.
- Mahner M., Bunge A. *Foundations of Biophilosophy*, Berlin, Springer, 1997.
- Stevens H. *Life Out of Sequence: A Data-Driven History of Bioinformatics*, Chicago, University of Chicago Press, 2013.
- Von Bertalanffy L. *General System Theory*, New York, George Braziller, 1973.
- Waterman M. S. *Introduction to Computational Biology: Maps, Sequences and Genomes*, Boca Raton, Chapman&Hall/CRC, 1995.
- Weinberg A. M. Impact of Large-Scale Science on the United States. *Science*, 1961, vol. 134, pp. 161–164.